



## ESTIMASI PARAMETER REGRESI QUANTIL DENGAN REGRESI *ROBUST LEAST TRIMMED SQUARE* (LTS) (Studi Kasus : Faktor -Faktor yang Mempengaruhi Penyebaran Penyakit Malaria di Indonesia)

<sup>1</sup>M. Y. MATDOAN<sup>1\*</sup>, B. W OTOK<sup>2</sup>, R. M. ATOK<sup>3</sup>

<sup>1</sup>Program Studi Statistika, FMIPA Universitas Pattimura,

<sup>23</sup>Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)

\*yahya.matdoan@fmipa.unpatti.ac.id

### ABSTRAK

Estimasi Metode OLS didasarkan pada distribusi normal, sehingga kurang tepat untuk menganalisis sejumlah data yang tidak simetris atau mengandung pencilan (*outlier*). Oleh karena itu, dikembangkan regresi quantil yang tidak terpengaruh oleh adanya *outlier*. Selain regresi quantil, regresi *robust* LTS juga dapat digunakan sebagai alternatif untuk menyelesaikan permasalahan data yang fluktuatif dan mengandung outlier. Data penyebaran penyakit malaria di Indonesia mengalami disparitas (kesenjangan) antar daerah dan mengandung *outlier*. Penelitian ini membandingkan regresi quantil dengan regresi *robust* LTS pada kasus faktor-faktor yang mempengaruhi penyakit malaria di Indonesia. Diperoleh hasil bahwa dengan menggunakan metode OLS diperoleh  $R^2$  sebesar 0,681 dan regresi *robust* LTS sebesar 0,732 serta regresi quantil dengan  $\tau = 0,95$  sebesar 0,832. sehingga diperoleh model terbaik yang diperoleh model regresi quantil. Lebih lanjut diperoleh hasil bahwa faktor utama penyebab penyebaran penyakit malaria di Indonesia yaitu faktor rumah layak huni, faktor penduduk miskin dan faktor tenaga dokter.

**Kata Kunci:** Regresi Quantil, Regresi *Robust* LTS, *Outlier*, Malaria

### ABSTRACT

OLS method estimation is based on a normal distribution, so it is not appropriate to analyze a number of data that are not symmetrical or contain outliers. Therefore, quantile regression is developed which is not affected by outlier. In addition to quantile regression, robust LTS regression can also be used as an alternative to solving data problems that are volatile and contain outliers. Aim of this study is to compare quantile regression and LTS robust regression in case of factors that contribute to spreading of malaria disease in Indonesia. The results obtain  $R^2$  of 0.681, 0.732, and 0.832 using OLS, LTS, and Quantile Regression ( $\tau = 0.95$ ) methods respectively. Conclusion of this study is quantile regression as the best model. Moreover the main factors that contribute to malaria disease spreading are the percentage of livable housing, the percentage of poor people, and doctors ratio.

**Keywords :** Quantile Regression, LTS Robust Regression, Outlier, Malaria

## 1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk menggambarkan hubungan antara dua atau lebih variabel, sehingga suatu variabel dapat diprediksikan dari variabel yang lain[2]. Variabel yang ditaksir nilainya disebut variabel respon ( $y$ ), sedangkan variabel penaksir disebut sebagai variabel prediktor ( $x$ ).

Pendekatan standar untuk mendapatkan nilai dugaan parameter dari model regresi linier adalah metode *Ordinary Least Square* (OLS). Dalam analisis regresi parametrik (khususnya regresi klasik), [7] menyatakan bahwa ada beberapa asumsi yang harus dipenuhi dalam melakukan estimasi, yaitu tidak terjadi multikolinieritas antar variabel prediktor, tidak terjadi autokorelasi dan heteroskedastisitas, serta residual berdistribusi normal. Jika semua asumsi tersebut terpenuhi, maka hasil estimasi dengan menggunakan metode OLS bersifat *Best Linear Unbiased Estimator* (BLUE)[8].

Estimasi parameter dengan menggunakan metode OLS diperoleh dengan meminimumkan jumlah kuadrat sisaan (residual) dan berdasarkan pada distribusi mean, dimana nilai mean menunjukkan ukuran pemusatan dari suatu distribusi data. Oleh sebab itu, pendekatan dengan menggunakan metode OLS tidak dapat merepresentasikan keseluruhan data[9]. Untuk mengatasi masalah tersebut, maka dikembangkanlah metode regresi kuantil. Regresi kuantil pertama kali diperkenalkan oleh Koenker dan Bassett (1978), Metode ini merupakan perluasan dari model regresi pada kuantil bersyarat[12]. Untuk mendapatkan estimasi parameter model regresi kuantil diperoleh dengan menggunakan metode pemrograman linier diantaranya algoritma *simpleks*, *interior-point*, dan *smoothing*[3]. Selain Regresi kuantil, regresi *robust Least Trimmed Square* merupakan metode yang dapat digunakan sebagai alternatif untuk menyelesaikan permasalahan data yang fluktuatif dan mengandung *outlier*.

Malaria adalah penyakit infeksi yang disebabkan oleh parasit *Plasmodium* yang hidup dan berkembang biak dalam sel darah merah manusia, ditularkan oleh nyamuk (*Anopheles*) betina dan dapat menyerang semua orang, baik laki-laki maupun perempuan pada semua golongan umur dari bayi, anak-anak maupun orang dewasa[16]. Penyakit malaria secara langsung dapat menyebabkan seseorang sering sakit, daya tahan tubuh rendah serta menyebabkan kematian.

Di Indonesia penyakit malaria ditemukan tersebar luas di berbagai daerah dengan derajat infeksi yang bervariasi[4]. Tingkat penyebaran penyakit malaria mencapai 6 persen dengan 5 Provinsi tertinggi yaitu Provinsi Papua, Nusa Tenggara Timur, Papua Barat, Sulawesi Tengah, dan Maluku. Kejadian tersebut disebabkan karena adanya permasalahan-permasalahan teknis seperti pembangunan yang tidak peduli terhadap kesehatan lingkungan, mobilitas penduduk dari daerah endemis malaria, wilayah terpencil dengan kondisi lingkungan yang tidak baik dan minimnya akses pelayanan kesehatan serta kondisi ekonomi masyarakat yang rendah[10]. Jumlah Penderita positif penyakit malaria di Indonesia pada tahun 2015 sebanyak 217.025 jiwa[1]. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia bahwa jumlah penderita penyakit malaria sudah mengalami penurunan. Namun, penurunan penderita penyakit malaria di Indonesia masih tetap harus diwaspadai[14], karena adanya disparitas (kesenjangan) data penyebaran penyakit malaria antar daerah di Indonesia[11]. Sehingga, diduga data penyebaran penyakit Malaria mengandung pencilan (*outlier*). Hal ini terutama pada daerah-daerah yang memiliki tingkat penyebaran penyakit malaria tinggi. Pencegahan penyebaran penyakit malaria di Indonesia sudah dilakukan oleh berbagai elemen terkait, namun belum maksimal. Oleh karena itu, perlu dilakukan pemodelan mengenai faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia.

Penelitian ini dilakukan dengan menggunakan metode Regresi Quantil dan Regresi *Robust Least Trimmed Square* (LTS) karena lebih sesuai untuk data penyebaran penyakit malaria di Indonesia.

## 2. Tinjauan Pustaka

### 2.1 Analisis Regresi

Analisis regresi pertama kali ditemukan oleh Sir Francis Galton (1822-1911). Analisis ini merupakan salah satu metode analisis yang digunakan untuk menyelidiki pola hubungan antara variabel respon (x) dengan variabel prediktor (y). Model umum persamaan regresi dapat dituliskan sebagai berikut :

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

dimana :

- $y_i$  : Variabel Respon pada pengamatan ke- $i$
- $f(x_i)$  : Kurva regresi pada pengamatan ke- $i$
- $\varepsilon_i$  : Residual (error) pada pengamatan ke- $i$

### 2.2 Pengujian *Outlier*

*Outlier* terjadi karena kombinasi unik yang terjadi dan nilai - nilai yang dihasilkan dari observasi tersebut sangat berbeda dari observasi-observasi lainnya[17]. Deteksi *outlier* dapat dilakukan dengan menggunakan *difference in fitted value (DFFITs)*[5].

$$DFFITs_i = \frac{\hat{y}_i - \hat{y}_{i,-i}}{\sqrt{S^2 h_{ii}}} \quad \text{untuk } i = 1, 2, \dots, n \quad (2)$$

### 2.3 *Ordinary Least Square (OLS)*

Metode *Ordinary Least Square (OLS)* pada prinsipnya adalah meminimumkan jumlah kuadrat error (residual). Selanjutnya ditunjukkan Penaksir parameter  $\hat{\beta}$  dari metode *Ordinary Least Square (OLS)* adalah sebagai berikut [7] :

$$\hat{\beta} = (X'X)^{-1} X'y \quad (3)$$

### 2.4 *Least Trimmed Square (LTS)*

*Least Trimmed Square (LTS)* merupakan metode estimasi parameter regresi *robust* untuk meminimumkan jumlah kuadrat sebanyak  $h$  residual[15].

$$\sum_{i=1}^h e_{(i)}^2 \quad (4)$$

Algoritma yang digunakan pada regresi *Robust Least Trimmed Square (LTS)* disebut sebagai *basic resampling algorithm*, seperti pertama kali dikenalkan oleh Roussew dan Leroy (1978) dalam PROGRESS (*Program for Robust Regression*)[18].

### 2.5 Regresi Quantil

Regresi quantil merupakan perluasan dari model regresi pada quantil bersyarat dimana distribusi quantil bersyarat dari variabel respon dinyatakan sebagai fungsi dari kovariat yang

diamati. Misalkan diberikan data  $\{x_{1i}, x_{2i}, \dots, x_{pi}, y_i\}, i=1, 2, \dots, n$  merupakan himpunan berpasangan variabel random yang terdistribusi secara independen dan identik dengan kuantil  $\tau \in (0,1)$ . Persamaan umum regresi kuantil linier yaitu sebagai berikut[6] :

$$y_i = \beta_{\tau 0} + \beta_{\tau 1}x_{1i} + \beta_{\tau 2}x_{2i} + \dots + \beta_{\tau p}x_{pi} + \varepsilon_{i\tau} \quad (5)$$

Selanjutnya persamaan (10) dapat ditulis dalam bentuk model linier berikut[12] :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \quad (6)$$

## 2.6 Estimasi Parameter Regresi Kuantil

Diberikan himpunan data berpasangan  $\{x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi}, y_i\}, i=1, 2, \dots, n, j=1, 2, \dots, p$  merupakan himpunan berpasangan variabel random yang terdistribusi secara independen dan identik dengan kuantil  $\tau \in (0,1)$ . Data tersebut memiliki fungsi distribusi peluang bersyarat yang didenifikasikan sebagai berikut[13] :

$$F(Y|x_i) = P(Y \leq y|x_i) \quad (7)$$

dan fungsi invers  $F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$  yang merupakan kuantil ke- $\tau$  dari variabel respon  $y$ . Jadi, persamaan umum regresi kuantil linier untuk kuantil bersyarat didenifikasikan sebagai berikut :

$$\begin{aligned} Q_y(\tau|x) &= F_y^{-1}(y|x_i) \\ &= \inf\{y : F_y(y|x) \geq \tau\} \end{aligned}$$

Sehingga model fungsi kuantil bersyarat yaitu sebagai berikut :

$$Q_y(\tau|x_i) = \mathbf{X}'\boldsymbol{\beta}(\tau), \tau \in (0,1) \quad (8)$$

Persamaan umum regresi kuantil linier dari variabel respon  $y$  dapat ditulis menjadi

$$\mathbf{y}_i = \beta_{\tau 0} + \beta_{\tau 1}x_{1i} + \beta_{\tau 2}x_{2i} + \dots + \beta_{\tau p}x_{pi} + \varepsilon_{i\tau} \quad (9)$$

Selanjutnya pada persamaan (14) dapat dibuat dalam bentuk matriks berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_{\tau 0} \\ \beta_{\tau 1} \\ \vdots \\ \beta_{\tau p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1\tau} \\ \varepsilon_{2\tau} \\ \vdots \\ \varepsilon_{n\tau} \end{bmatrix}$$

Sehingga diperoleh bentuk model linier berikut :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \quad (10)$$

Dalam regresi klasik (metode OLS) dengan prinsip meminimumkan jumlah kuadrat residual untuk mencari penduga  $\hat{\boldsymbol{\beta}}$ , begitu halnya dengan regresi kuantil. Regresi kuantil ke- $\tau$  dari  $F_y$  dapat diperoleh dari persamaan (11) yaitu sebagai berikut :

$$E\rho(Y - \hat{y}) = \int_{-\infty}^{\hat{y}} (1-\tau)(y - \hat{y})f(y)dy + \int_{\hat{y}}^{\infty} \tau(y - \hat{y})f(y)dy \quad (11)$$

Selanjutnya pada persamaan (11) diminimumkan menjadi nol :

$$\begin{aligned} \frac{\partial}{\partial \hat{y}} E[\rho_{\tau}(Y - \hat{y})] &= 0 \\ \frac{\partial}{\partial \hat{y}} \int_{-\infty}^{\hat{y}} (1-\tau)(y - \hat{y}) f(y) dy + \int_{\hat{y}}^{\infty} \tau(y - \hat{y}) f(y) dy &= 0 \\ \frac{\partial}{\partial \hat{y}} \left[ (1-\tau) \int_{-\infty}^{\hat{y}} (y - \hat{y}) f(y) dy + \tau \int_{\hat{y}}^{\infty} (y - \hat{y}) f(y) dy \right] &= 0 \\ (1-\tau) \left[ (y - \hat{y}) f(y) \Big|_{-\infty}^{\hat{y}} + \int_{-\infty}^{\hat{y}} \frac{\partial}{\partial \hat{y}} (y - \hat{y}) f(y) dy \right] + \\ \tau \left[ (y - \hat{y}) f(y) \Big|_{\hat{y}}^{\infty} + \int_{\hat{y}}^{\infty} \frac{\partial}{\partial \hat{y}} (y - \hat{y}) f(y) dy \right] &= 0 \\ (1-\tau) \left[ (y - \hat{y}) f(y) \Big|_{y=\hat{y}} + \int_{-\infty}^{\hat{y}} f(y) dy \right] + \\ \tau \left[ (y - \hat{y}) f(y) \Big|_{y=\hat{y}} + \int_{\hat{y}}^{\infty} f(y) dy \right] &= 0 \\ (1-\tau)[0 + F_Y(\hat{y})] + \tau[0 - (1 - F_Y(\hat{y}))] &= 0 \\ (1-\tau)F_Y(\hat{y}) - \tau(1 - F_Y(\hat{y})) &= 0 \\ (1-\tau)F_Y(\hat{y}) - \tau - \tau F_Y(\hat{y}) &= 0 \\ F_Y(\hat{y}) - \tau &= 0 \end{aligned}$$

maka diperoleh

$$F_Y(\hat{y}) = \tau$$

sehingga quantil ke- $\tau$  merupakan solusi dari  $F_Y$ .

Regresi quantil ke- $\tau$  diperoleh dengan meminimumkan jumlah nilai mutlak dari error dengan pembobot ( $\tau$ ) untuk error positif dan pembobot  $(1-\tau)$  untuk error negatif. sehingga diperoleh solusi persamaan berikut :

$$\hat{\beta}(\tau) = \min_{\beta \in R^{p+1}} \tau \sum_{y \geq x} |y - \mathbf{X}^T \beta| + (1-\tau) \sum_{y < x} |y - \mathbf{X}^T \beta| \quad (12)$$

atau dapat ditulis seperti persamaan (18) berikut :

$$\hat{\beta}(\tau) = \min_{\beta \in R^{p+1}} \sum_{i=1}^n \rho_{\tau}(u_i) \quad (13)$$

Dalam estimasi OLS dari suatu model linier terhadap  $y$  diperoleh dengan meminimumkan jumlah kuadrat error. Sedangkan dalam estimasi regresi quantil dari suatu model linier terhadap  $y$  diperoleh dengan meminimumkan nilai *loss function* yang tak simetris yaitu meminimumkan nilai harapan  $\rho_{\tau}(u)$ . Selanjutnya akan dibuktikan bahwa *loss function* berbentuk asimetris dengan penjelasan berikut :

diberikan

$$\begin{aligned} \rho_{\tau} &= [\tau I(u \geq 0) + (1-\tau)I((u < 0))] |u| \\ &= [\tau - I(u < 0)] u \end{aligned}$$

dengan

$$I(u \geq 0) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases} \quad \text{dan} \quad |u| = \begin{cases} u, & u \geq 0 \\ -u, & u < 0 \end{cases}$$

dimana

$u$  : Error dari penduga parameter

$I(u)$  : Fungsi indikator yang telah didefinisikan

Sehingga dapat dibuktikan

$$\rho_\tau = \begin{cases} \tau u, & u \geq 0 \\ (\tau-1)u, & u < 0 \end{cases}$$

a. Untuk  $u \geq 0$

$$\begin{aligned} \rho_\tau &= [\tau I(u \geq 0) + (1-\tau)I(u < 0)]|u| \\ &= [\tau I(u \geq 0) + (1-\tau)I(u < 0)]u \\ &= [\tau + (1-\tau)I(u < 0)]u \\ &= [\tau + I(u < 0) - \tau I(u < 0)]u \\ &= [\tau + (1 - I(u \geq 0)) - \tau(1 - I(u \geq 0))]u \\ &= [\tau + (1-1) - \tau(1-1)]u \\ &= \tau u \end{aligned}$$

dapat juga diselesaikan dengan cara seperti berikut :

$$\begin{aligned} \rho_\tau &= [\tau - I(u < 0)]u \\ &= [\tau - (1 - I(u \geq 0))]u \\ &= [\tau - (1-1)]u \\ &= \tau u \end{aligned}$$

b. Untuk  $u < 0$

$$\begin{aligned} \rho_\tau &= [\tau I(u \geq 0) + (1-\tau)I(u < 0)]|u| \\ &= [\tau I(u \geq 0) + (1-\tau)I(u < 0)](-u) \\ &= [\tau 0 + (1-\tau)I(u < 0)](-u) \\ &= [(\tau-1)I(u < 0)]u \\ &= [(\tau-1)(1 - I(u < 0))]u \\ &= [(\tau-1)(1-0)]u \\ &= (\tau-1)u \end{aligned}$$

dapat juga diselesaikan dengan cara seperti berikut :

$$\begin{aligned} \rho_\tau &= [\tau - I(u < 0)]u \\ &= [\tau - (1 - I(u \geq 0))]u \\ &= [\tau - (1-0)]u \\ &= (\tau-1)u \end{aligned}$$

Sehingga menjadi

$$\rho_\tau = [\tau I(u \geq 0) + (1-\tau)I(u < 0)]|u| = [\tau - I(u < 0)]u, \forall u$$

dengan

$$\rho_\tau = \begin{cases} \tau u & , \text{ jika } u \geq 0 \\ (\tau-1)u & , \text{ jika } u < 0 \end{cases}$$

Solusi dari persamaan (12) tidak dapat diperoleh secara analitik, tetapi secara numerik. Salah satu metode numerik yaitu algoritma simpleks. Metode algoritma simpleks telah dikembangkan oleh Barrodale dan Robert pada tahun 1974. Metode algoritma tersebut memberikan solusi permasalahan program linier yang melibatkan banyak variabel - variabel keputusan dengan bantuan komputasi.

Adapun langkah-langkah proses optimasi metode simpleks adalah sebagai berikut :  
Diberikan bentuk umum regresi kuantil

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau)$$

Langkah awal digunakan kuantil  $\tau = 0.05$  untuk mencari nilai estimasi  $\hat{\boldsymbol{\beta}}$ , maka harus meminimumkan fungsi

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n |\mathbf{y} - \mathbf{X}\boldsymbol{\beta}|$$

Diberikan  $[x]_+$ , non negatif bagian dari  $x$ , yaitu  $\mathbf{s}_1 = [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]_+$  dan  $\mathbf{s}_2 = [\mathbf{X}\boldsymbol{\beta} - \mathbf{y}]_+$  maka masalah linier dapat ditulis

$$\min_{\boldsymbol{\beta}} \{ \mathbf{1}'\mathbf{s}_1 + \mathbf{1}'\mathbf{s}_2 \mid \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{s}_1 - \mathbf{s}_2, \{\mathbf{s}_1, \mathbf{s}_2\} \in \mathbf{R}_+^n \}$$

Selanjutnya diberikan

$$\mathbf{B} = [\mathbf{X} \ -\mathbf{X}\mathbf{I} \ -\mathbf{I}] \text{ dan } \boldsymbol{\psi} = \begin{bmatrix} [\boldsymbol{\beta}] \\ [-\boldsymbol{\beta}] \\ [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}] \\ [\mathbf{X}\boldsymbol{\beta} - \mathbf{y}] \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 0_{[p]} \\ 0_{[p]} \\ \mathbf{1}_{[n]} \\ \mathbf{1}_{[n]} \end{bmatrix}$$

Seperti permasalahan pemrograman linier standar, maka formulasinya dapat ditulis  $\min \mathbf{d}'\boldsymbol{\psi}$  dimana untuk  $\mathbf{B}\boldsymbol{\psi} = \mathbf{y}$ ,  $\tau \geq 0$ . Oleh karena itu parameter dual dapat ditulis  $\max_d \mathbf{y}'\mathbf{z}$  dimana untuk  $\mathbf{B}'\mathbf{z} \leq \mathbf{d}$ . Mengingat hasil utama program linier yaitu solusi dari masalah minimasi tersebut diselesaikan dengan metode simpleks. Secara sederhana masalah diatas dapat dirumuskan menjadi  $\max_{\mathbf{z}} \{ \mathbf{y}'\mathbf{z} \mid \mathbf{X}'\mathbf{z} = \mathbf{0}, \mathbf{z} \in [-1, +1]^n \}$ , untuk  $\mathbf{X}'\mathbf{z} = 0$  dapat ditransformasikan dengan cara dikalikan dengan  $\frac{1}{2}$  dan  $\frac{1}{2}\mathbf{X}'\mathbf{z} + \frac{1}{2}\mathbf{X}'\mathbf{1} = \frac{1}{2}\mathbf{X}'\mathbf{1}$ , untuk yang ditambah  $\frac{1}{2}\mathbf{X}'\mathbf{1}$  dengan kata lain dapat ditulis seperti pada persamaan berikut :

$$\mathbf{X}'\left(\frac{1}{2}\mathbf{z} + \frac{1}{2}\mathbf{1}\right) = \frac{1}{2}\mathbf{X}'\mathbf{1} \quad (14)$$

Jika  $\frac{1}{2}\mathbf{z} + \frac{1}{2}\mathbf{1} = \boldsymbol{\eta}$  dan  $\frac{1}{2}\mathbf{X}'\mathbf{1} = \mathbf{b}$ , maka persamaan dapat ditulis  $\mathbf{X}'\boldsymbol{\eta} = \mathbf{b}$

dan  $\max_{\mathbf{J}} \{ \mathbf{y}'\mathbf{J} \mid \mathbf{X}'\mathbf{J} = \mathbf{b}, \mathbf{J} \in [0, 1]^n \}$ . Nilai  $\frac{1}{2}$  pada persamaan (14) di atas, merupakan kunci utama generalisasi untuk kondisi kuantil yang lain. Bahkan masalah minimalisasi untuk median bersyarat dapat dipakai untuk kuantil bersyarat ke- $\tau$  sehingga diperoleh

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho_{\tau}(\mathbf{y} - \mathbf{X}'\boldsymbol{\beta}(\tau))$$

cara yang sama diperoleh  $\max_{\mathbf{z}} \{ \mathbf{y}'\mathbf{z} \mid \mathbf{X}'\mathbf{z} = (1-\tau)\mathbf{X}'\mathbf{1}, \mathbf{z} \in [0, 1]^n \}$  dimana  $(1-\tau)$  memiliki peran yang sama seperti  $\frac{1}{2}$  pada rumus di median.

### 3. Metodologi Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Kementerian Kesehatan Republik Indonesia dan Badan Pusat Statistika (BPS) pada tahun 2015. Variabel yang digunakan dalam penelitian ini terdiri dari respon dan prediktor. Variabel respon yang digunakan yaitu Annual parasite incidence (API) malaria (Y) dan variabel prediktor yang digunakan yaitu presentase akses sanitasi layak (X<sub>1</sub>), presentase rumah layak huni (X<sub>2</sub>), presentase penduduk miskin (X<sub>3</sub>), rasio tenaga dokter (X<sub>4</sub>), presentase akses air bersih (X<sub>5</sub>) dan rasio puskesmas (X<sub>6</sub>).

## 4. Hasil dan Pembahasan

### 4.1 Analisis Deskriptif dan Identifikasi Data

Hasil statistik deskriptif pada kasus penyakit malaria dan faktor-faktor yang mempengaruhi penyebaran penyakit malaria disajikan pada Tabel 1 berikut.

**Tabel 1.** Statistik Deskriptif

Deskriptif	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
Mean	2,63	59	89,8	11,7	20,1	68,6	1,86
Min	0	23,9	49,2	3,6	10,9	41,0	0,58
Maks	31,9	89,2	99,2	28,4	39,1	93,4	5,20
Range	31,9	65,3	49,9	24,7	28,2	52,3	4,62
Std Dev	7,51	15,3	10,4	6,1	6,8	11,0	1,07
Variansi	56,4	236	108,3	38,2	46,4	121,7	1,15

### 4.2. Pengujian *Outlier*

Pengujian data *outlier* dengan menggunakan metode *DFFITs*. Pengamatan dinyatakan sebagai *outlier* jika  $|DFFITs| > 2\sqrt{(k+1)/n} = 0,907$ .

**Tabel 2.** Pengujian *Outlier*

Propinsi	<i>DFFITs</i>
Nusa Tenggara Timur	1,69655
Sulawesi Tenggara	1,28412
Papua Barat	6,41416
Papua	4,55268

Berdasarkan Tabel 2, dapat dilihat bahwa terdapat empat Propinsi yang dinyatakan sebagai *outlier* yaitu Propinsi Nusa Tenggara Timur (NTT), Propinsi Sulawesi Tenggara, Propinsi Papua Barat dan Propinsi Papua. Apabila analisis regresi dilakukan dengan menggunakan metode OLS, akan mengakibatkan varians dari penduga parameter besar, sehingga tidak efisien. Oleh karena itu, digunakan regresi kuantil dan regresi *robust* LTS untuk menentukan parameter dari faktor-faktor yang mempengaruhi peningkatan presentase penderita penyakit malaria di Indonesia.

### 4.3. Estimasi Parameter

#### 4.3.1. Estimasi Parameter dengan Menggunakan Metode *Ordinary Least Square* (OLS)

Hasil Estimasi metode *Ordinary Least Square* (OLS) dapat dilihat pada Tabel 3 berikut.

**Tabel 3.** Estimasi Parameter *Ordinary Least Square* (OLS)

Parameter	Estimasi	P- Value	R <sup>2</sup>
$\beta_0$	2,211	0,865	
$\beta_1$	0,095	0,314	
$\beta_2$	-0,255	0,071	
$\beta_3$	0,451	0,028	0,681
$\beta_4$	0,024	0,870	
$\beta_5$	0,091	0,417	
$\beta_6$	3,073	0,008	

Berdasarkan Tabel 3, dapat dilihat bahwa terdapat dua variabel independen yang berpengaruh signifikan terhadap banyaknya *Annual Parasite Incidence* (API) Malaria di Indonesia yaitu variabel presentase penduduk miskin ( $X_3$ ) dan variabel rasio puskesmas ( $X_6$ ). Hasil Estimasi tersebut dapat dimodelkan sebagai berikut :

$$\hat{Y} = 2,211 + 0,095X_1 - 0,255X_2 + 0,451X_3 + 0,024X_4 + 0,091X_5 + 3,073X_6$$

#### 4.3.2. Estimasi Parameter dengan Metode Regresi *Robust LTS*

Metode regresi *robust LTS* dapat digunakan untuk mengatasi pengamatan *outlier*.

**Tabel 4.** Estimasi Parameter dengan *Robust LTS*

Parameter	Estimasi	P- Value	R <sup>2</sup>
$\beta_0$	-1,324	0,231	
$\beta_1$	-0,001	0,774	
$\beta_2$	0,018	0,177	
$\beta_3$	0,035	0,000	0,732
$\beta_4$	0,029	0,000	
$\beta_5$	-0,014	0,022	
$\beta_6$	0,069	0,204	

Berdasarkan Tabel 4, dapat dilihat bahwa terdapat tiga variabel independen yang berpengaruh secara signifikan terhadap banyaknya jumlah penderita API malaria di Indonesia yaitu variabel presentase kemiskinan ( $X_3$ ), variabel rasio tenaga kesehatan ( $X_4$ ) dan variabel presentase akses air bersih ( $X_5$ ). Pemodelan metode regresi *robust LTS* sebagai berikut :

$$\hat{Y} = -1,324 - 0,001X_1 + 0,018X_2 + 0,035X_3 + 0,029X_4 - 0,014X_5 + 0,069X_6$$

#### 4.3.3. Estimasi Parameter dengan Metode Regresi *Quantil*

Berdasarkan statistika deskriptif menunjukkan bahwa terdapat *outlier* dan tidak semetris pada data. sehingga, digunakan analisis regresi *quantil*. Langkah pertama dalam analisis ini yaitu memilih nilai *quantil* yang ingin diestimasi dengan *quantil*  $\tau = 0.95$ .

**Tabel 5.** Estimasi Parameter Regresi Quantil

Quantil	Parameter	Estimasi	P- Value	R <sup>2</sup>
Q <sub>(0,95)</sub>	$\beta_0$	11,934	0,031	0,83
	$\beta_1$	0,170	0,000	
	$\beta_2$	-0,394	0	
	$\beta_3$	0,407	0,000	
	$\beta_4$	-0,275	0,000	
	$\beta_5$	0,161	0,001	
	$\beta_6$	5,704	0	

Berdasarkan hasil estimasi pada Tabel 5, Diperoleh model regresi quantil sebagai berikut :

$$Q_{0,95}(y|x) = 11,934 + 0,170X_1 - 0,3947X_2 + 0,407X_3 - 0,275X_4 + 0,161X_5 + 5,704X_6$$

Berdasarkan Tabel 5, dapat diketahui bahwa dengan menggunakan taraf signifikan ( $\alpha = 5\%$ ) pada quantil  $Q_{(0,95)}$ , maka dapat disimpulkan bahwa variabel yang berpengaruh secara signifikan terhadap peningkatan *Annual Parasite Incidence* (API) malaria di Indonesia yaitu variabel presentase akses sanitasi layak ( $X_1$ ), variabel presentase rumah layak huni ( $X_2$ ), variabel presentase kemiskinan ( $X_3$ ), variabel rasio banyaknya tenaga dokter ( $X_4$ ) dan variabel presentase akses air bersih ( $X_5$ ) serta variabel rasio banyaknya puskesmas ( $X_6$ ). Pada quantil  $Q_{(0,95)}$ , diperoleh *koefisien determinasi* atau  $R^2$  sebanyak 0,83. Hal ini menunjukkan bahwa 83% kasus penyebaran penyakit malaria di Indonesia mampu dijelaskan oleh model, sedangkan sisanya sebesar 17% dijelaskan oleh peubah lain di luar model.

#### 4.4. Pemilihan Model Terbaik

Setelah mendapatkan model regresi quantil, maka langkah selanjutnya adalah memilih model terbaik dengan menggunakan nilai *koefisien determinasi* ( $R^2$ ) yang paling tinggi.

**Tabel 6.** Koefisien Determinasi ( $R^2$ )

Nilai Quantil	( $R^2$ )
Q <sub>(0,95)</sub>	0,832*
OLS	0,681
Robust LTS	0,732

Berdasarkan Tabel 6. dapat dilihat bahwa pada nilai quantil  $Q_{(0,95)}$ , memiliki nilai *koefisien determinasi*  $R^2$  yaitu sebesar (0,832) lebih besar jika dibandingkan dengan metode *Ordinary Least Square* (OLS) dengan  $R^2$  sebesar (0,681) serta metode regresi *robust LTS* dengan  $R^2$  sebesar (0,732). Hal ini menunjukkan bahwa model regresi pada quantil  $Q_{(0,95)}$  merupakan model terbaik.

Berdasarkan model terbaik yaitu model regresi quantil  $Q_{(0,95)}$ , menunjukkan bahwa variabel presentase akses sanitasi layak ( $X_1$ ) berpengaruh positif dan signifikan terhadap *annual parasite incidence* (API) malaria. Pada quantil  $Q_{(0,95)}$  menunjukkan bahwa koefisien variabel persentase akses sanitasi layak ( $X_1$ ) sebanyak 0,17048. Hal ini berarti bahwa daerah dengan akses sanitasi layak tinggi cenderung memiliki tingkat kerentanan terhadap penyakit malaria. Hal ini tidak sesuai dengan hipotesis yang diharapkan yaitu akses sanitasi layak dapat menurunkan morbiditas penyakit malaria. Ketidaksiharian hasil ini, diduga disebabkan karena kurang bervariasinya data morbiditas atau API penderita penyakit malaria.

Selanjutnya pada quantil  $Q_{(0,95)}$  menunjukkan bahwa variabel presentase rumah layak huni ( $X_2$ ), berpengaruh negatif dan signifikan terhadap API malaria. Koefisien presentase rumah layak huni ( $X_2$ ) sebanyak minus 0,39447. Artinya jika persentase rumah layak huni di suatu daerah naik sebanyak 1 persen, maka akan menurunkan tingkat *Annual Parasite Incidence* (API) malaria sebanyak 0,39 per 1000 penduduk. Hal ini sesuai dengan hipotesis yang diharapkan bahwa kondisi rumah layak huni dapat mengurangi *Annual Parasite Incidence* (API) malaria. Kondisi rumah dengan dinding yang terbuat dari kayu memungkinkan lebih banyak lubang untuk masuknya nyamuk. Demikian pula lantai rumah dari tanah menyebabkan kondisi rumah menjadi lembab sehingga menjadi tempat yang ideal bagi nyamuk.

Selanjutnya pada quantil  $Q_{(0,95)}$  menunjukkan bahwa variabel presentase kemiskinan ( $X_3$ ) berpengaruh positif dan signifikan terhadap *Annual Parasite Incidence* (API) malaria. Koefisien presentase kemiskinan ( $X_3$ ) sebanyak 0,40715. Artinya jika persentase kemiskinan di suatu daerah naik sebanyak 1 persen, maka akan meningkatkan tingkat *Annual Parasite Incidence* (API) malaria sebanyak 0,40 per 1000 penduduk. Hal ini menunjukkan bahwa ada kecenderungan daerah dengan presentase kemiskinan relatif tinggi akan diikuti tingkat *Annual Parasite Incidence* (API) malaria tinggi pula. Hal ini sesuai dengan hipotesis yang diharapkan bahwa tingkat kemiskinan dapat meningkatkan morbiditas malaria. Hubungan kemiskinan dengan tingkat *Annual Parasite Incidence* (API) malaria dapat dilihat dari turunan variabel kemiskinan, misalnya penggunaan kelambu dan pekerjaan. Pada banyak penelitian disebutkan bahwa penduduk yang menggunakan kelambu resiko terkena malaria lebih rendah dibanding yang tidak memakai. Demikian pula ada beberapa pekerjaan beresiko seperti nelayan dan pertanian yang lebih rentan terkena malaria dibanding jenis pekerjaan lainnya.

Selanjutnya pada quantil  $Q_{(0,95)}$  menunjukkan bahwa variabel rasio tenaga dokter ( $X_4$ ) berpengaruh negatif dan signifikan terhadap *Annual Parasite Incidence* (API) malaria. Koefisien variabel rasio tenaga dokter ( $X_4$ ) sebanyak minus 0,27588. Artinya jika rasio tenaga dokter di suatu daerah naik sebanyak 1 satuan, maka akan menurunkan tingkat *Annual Parasite Incidence* (API) malaria sebanyak 0,27 per 1000 penduduk. Artinya ada kecenderungan daerah dengan rasio tenaga dokter relatif tinggi akan diikuti tingkat *Annual Parasite Incidence* (API) malaria yang relatif lebih rendah. Hal ini sesuai dengan hipotesis yang diharapkan bahwa tenaga kesehatan dapat menurunkan tingkat *Annual Parasite Incidence* (API) malaria. Hubungan tenaga kesehatan terhadap kerentanan penyakit malaria dapat dijelaskan melalui kegiatan pelayanan kesehatan. Tenaga kesehatan dapat melakukan penyuluhan, penyemprotan, dan pengobatan sehingga dapat mempengaruhi *Annual Parasite Incidence* (API) malaria di suatu daerah.

Selanjutnya pada quantil  $Q_{(0,95)}$  menunjukkan bahwa variabel presentase akses air bersih ( $X_5$ ) berpengaruh positif dan signifikan terhadap *Annual Parasite Incidence* (API) malaria. Koefisien variabel presentase akses air bersih ( $X_5$ ) sebanyak 0,16171. Artinya daerah dengan akses air bersih tinggi cenderung memiliki tingkat kerentanan terhadap penyakit malaria. Hal ini tidak sesuai dengan hipotesis yang diharapkan yaitu akses air bersih dapat menurunkan tingkat *Annual Parasite Incidence* (API) malaria. Ketidaksesuaian hasil ini diduga disebabkan karena kurang bervariasinya data *Annual Parasite Incidence* (API) malaria. Hal lainnya diduga karena akses air bersih pada daerah yang rendah tingkat *Annual Parasite Incidence* (API) malariannya adalah air leding atau air mengalir lainnya. Nyamuk penyebab penyakit malaria biasanya berada pada air tergenang seperti seperti air sumur yang tidak terlindungi dan memiliki presentase penggunaannya rendah di Indonesia. Sehingga, dampak dari pengaruh air bersih terhadap tingkat *Annual Parasite Incidence* (API) malaria menjadi tidak bermakna.

Selanjutnya pada kuantil  $Q_{(0,95)}$  menunjukkan bahwa variabel rasio banyaknya puskesmas ( $X_6$ ) berpengaruh positif dan signifikan terhadap *Annual Parasite Incidence* (API) malaria. Koefisien rasio banyaknya puskesmas ( $X_6$ ) sebanyak 5,7049. Artinya bahwa ada kecenderungan daerah dengan rasio banyaknya puskesmas relatif tinggi, akan diikuti dengan tingkat *Annual Parasite Incidence* (API) malaria yang relatif tinggi pula. Hal ini tidak sesuai dengan hipotesis yang diharapkan bahwa banyaknya puskesmas dapat menekan tingkat *Annual Parasite Incidence* (API) malaria. Ketidaksesuaian hubungan ini, diduga disebabkan karena jumlah puskesmas yang banyak tidak menjamin akses ke sarana kesehatan yang lebih baik bagi masyarakat. Hal ini, dimungkinkan pada kawasan Timur Indonesia, sarana puskesmas yang hanya terletak di Ibu Kota Kabupaten dan beberapa Ibu Kota Kecamatan sehingga persebarannya tidak merata. Selain itu, jumlah puskesmas yang banyak namun tidak diikuti dengan pelayanan yang maksimal akan menyebabkan ketimpangan dalam akses sarana kesehatan bagi masyarakat.

## 5. Kesimpulan

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan dari penelitian ini adalah sebagai berikut :

1. Hasil perbandingan metode estimasi yang diperoleh yaitu metode Regresi Kuantil lebih baik dibandingkan dengan metode Regresi *Robust Least Trimmed Square* (LTS) dan metode *Ordinary Least Square* (OLS) pada kasus faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia. Hal ini, ditunjukkan dengan nilai  $R^2$  dari model Regresi Kuantil sebesar 0,832, model Regresi *Robust Least Trimmed Square* (LTS) sebesar 0,732 dan model *Ordinary Least Square* (OLS) sebesar 0,681.
2. Faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia dengan menggunakan model terbaik, pada model regresi kuantil dengan  $\tau=0,95$  adalah faktor rumah layak huni, faktor penduduk miskin dan faktor tenaga dokter. Oleh karena itu, Pemerintah beserta seluruh elemen masyarakat perlu mengevaluasi dan melakukan proses pencegahan agar terhindar dari penyebaran penyakit malaria.

## Daftar Pustaka

- [1] Badan Pusat Statistika, Hasil Survei Sosial Ekonomi Nasional Tahun 2014, Jakarta, 2014.
- [2] I. Budiantara, "Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter.," in *Seminar Nasional FMIPA Universitas Pendidikan Ganesha.*, Universitas Pendidikan Ganesha., 2011.
- [3] C. d. W. Chen, "Computational Issues for Quantile Regression," *The Indian Journal of Statistics*, vol. 67, pp. 399-417, 2005.
- [4] M. d. M. Datukramat, Menganalisis Hubungan Anatara Sanitasi Lingkungan Dengan Kejadian Malaria Pada Murid Sekolah Dasar Di Kabupaten Bolaang Mongondow Utara, Manado: journal keperawatan (e-Kp) Fakultas Kedokteran Universitas Sam Ratulangi, 2013.
- [5] I. Ghozali, Aplikasi analisis Multivariat dengan program SPSS, Semarang : Edisi ketujuh, 2013.

- [6] S. C. d. K. K. Gob, "Nonstandard Quantile-Regression Inference," *Econometric Theory*, vol. 25, pp. 1415-1432, 2009.
- [7] D. N. Gujarati, *Basic Econometrics*, Newyork: 4th Edition. New York: McGrahill. Co., 2004.
- [8] W. Hardle, *Applied Nonparametric Regression*, New York: Cambridge University Press, 1990.
- [9] T. J. d. T. R. J. Hastie, *Generalized Additive Models*, New York. London: Chapman and Hall, 1990.
- [10] I. d. Yuyun, "Analisis faktor-faktor yang mempengaruhi angka kejadian penyakit tropis malaria dengan metode klasifikasi pohon (Studi kasus Kabupaten Sorong Selatan, Provinsi Papua Barat).," *Jurnal, Institut Teknologi Sepuluh Nopember*, 2010.
- [11] K. K. R. Indonesia, *Profil kesehatan Indonesia*, Jakarta: Pusat data dan Informasi Kementerian Kesehatan, 2014.
- [12] R. d. M. J. A. F. Koenker, "Goodness of fit and Related Inference Process for Quantile Regression," *Journal of the American statistical Association*, vol. 94, no. 448, pp. 1296-1310, 1999.
- [13] R. Koenker, *Ekonometric society monographs "Quantile Regression"*, New York: Cambridge University, 2005.
- [14] Z. L. Lin W, "Optimal smoothing in nonparametric conditional quantile derivative function estimation," *Journal of Econometrics Science Direct*, vol. 188, pp. 502-513., 2015.
- [15] A. I. A. d. M. R. Mirontoneng, "Analisis faktor-faktor yang berhubungan kejadian malaria pada anak di wilayah kerja PKM Tona kecamatan tahuna kabupaten Sangihe," *Program Studi Ilmu Keperawatan Fakultas kedokteran Universitas Sam Ratula*, 2014.
- [16] R. H. Myers, *Classical and Modern Regression with Application*, Boston: PWS, 1990.
- [17] P. A. Susilowati, "Analisis regresi pada prevalensi malaria di Provinsi Maluku Utara, Maluku, Papua Barat dan Papua dengan faktor yang mempengaruhinya," *Skripsi, Jurusan Statistika Institut Teknologi Sepuluh Nopember*, 2013.
- [18] M. Werner, *Identification of Multivariate Outliers in Large Data Sets*, University of Colorado at Denver, 2003.